BIOLOGIJA. 2005. Nr. 1. P. 1–5

Evaluation of some S/MAR sequence properties on DNA sequences involved in tightly bound DNA protein complexes of murine cells

D. Labeikytë*,
L. Bagdonienë,
V. Jonuðienë,
S. Sasnauskienë,
B. Juodka

Department of Biochemistry and Biophysics, Vilnius University M. K. Èiurlionio 22, LT-03101 Vilnius, Lithuania

Abbreviations: MEL – Friend erythroleukemia cells; TBD – eukaryotic non-histone proteins tightly bound to DNA; S/MARs – conserved DNA sequences associated with matrix and/or scaffold

Here we report the DNA sequence-based characterization of DNA fragments involved in stable complexes with proteins in mouse chromatin. Computer analysis of nine independently isolated and sequenced clones containing such sequences revealed some similar properties known to comprise S/MAR sequences. These characteristics (ori signals, TA, CA, TG richness, curved DNA, topoisomerase II binding sites) suggest that DNAs presented in tightly DNA-bound complexes have a certain characteristic sequence and conformation properties typical of S/MARs.

Key words: chromosomal DNA loops, nuclear matrix DNA, S/MAR sequences

INTRODUCTION

One much debated possibility is that the structure and function of the nucleus are determined by a proteinaceous structure called nuclear matrix [1, 2]. Attachment of chromatin to the nuclear matrix occurs via specialized DNA elements that have been found in all eukaryotic organisms studied. Termed S/MARs, these DNA elements are bound by nuclear matrix in an evolutionary conserved manner, presumably because of conserved binding proteins present in these matrices [2]. S/MARs organize the eukaryotic genome into topologically independent transcriptional domains [3], increase the activity of nearby enhancers [4], may influence the expression of integrated tumor virus genomes [5], play a fundamental role in carcinogenesis [6].

The persistent fraction of polypeptides is copurified with eukaryotic DNA which points to very tight (maybe covalent) complexes between nuclear proteins and DNA [7, 8]. Such tightly bound (TBD) complexes are of interest with respect to their potential involvement in the topological organization and/or function of genomic DNA. It has been shown that the protein component in such complexes is stable and cannot be removed after repeated ethanol precipitations and incubation in different ionic strength buffers, by exposure to 0.2 M guanidine · HCl, 4M urea, 2% sarcosyl, 2% SDS, 0.1 M boric acid, 0.02 M phenanthroline, 0.01 M EDTA, 0.1 N NaOH and high concentrations of salt [9]. Attempts to characterize the DNA component of such complexes have been reported, but due to the different isolation procedures used, varying results were obtained [10, 11]. It has been suggested that the most stable DNA-bound proteins (resisting salt, urea, detergent treatment and phenol extractions) may be involved in the formation of another type of loops, namely permanently attached constitutive loops [12], and that such DNA–protein complexes could be transmitted to the daughter cells as stable structural units [11]. Therefore elucidation of the nature of DNA sequences involved in stable (maybe covalent) complexes with nuclear proteins is an important step toward understanding their function.

Our present goal is to determine the sequence-based nature of DNAs separated from the tightly bound DNA–protein complexes of murine Friend erythroleukemia cells and to compare them with the properties of known S/MAR sequences.

*) Corresponding author: Danutë Labeikytë, Vilniaus Universitetas, Gamtos mokslø fakultetas, Biochemijos ir biofizikos katedra, M. K. Èiurlionio 21, LT-03101 Vilnius, Lithuania

E-mail: danute.labeikyte@gf.vu.lt

MATERIALS AND METHODS

Materials

[α-^{32}P]dATP (3000 Ci/mmol : NEN) was from Amersham; Ligation and Sequencing kits were from Fermentas; Other chemicals were of possibly highest quality.

Cells culture

Friend erythroleukemia cells, clone F4N were grown in RPMI-1640 medium (Sigma) supplemented with 10% (v/v) heat-inactivated fetal calf serum. The cells were kept in logarithmic grown phase (1^x10^5–2^x10^6 cells /ml) with appropriate dilutions every 3–4 days. Cell viability, evaluated by the Trypan blue dye exclusion test, was never lower than 95%.

Cell lysis and DNA salting-out procedure were described in detail elsewhere [13].

Separation of DNA protected by co-isolated polypeptides. Aliquots containing 4 mg of DNA then were precipitated with 2 vol ethanol. DNA pellets were collected, dried and dissolved in 50 µl DNase I buffer (50 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$). DNase I (Pharmacia, RNase free) (10 u/mg DNA) were added and, after overnight incubation at room temperature under dialysis conditions against DNase I buffer, DNA–polypeptide complexes released by DNase I digestion were subjected to 1% agarose gel electrophoresis. Material from 1% agarose gel start was recovered and treated with proteinase K (1mg/ml) in TE0.1 (10 mM Tris-HCl, pH 7.5; 0.1 mM EDTA) containing 0.5% SDS (3 h, 37 °C). Then the digest was twice phenolyzed, precipitated with ethanol, and the resulting DNA was subjected to standard cloning and sequencing procedures. The residual phenolyzed and ethanol-precipitated DNA were blunt-ended and ligated with the pUC19/*Sma*I vector. Transformation of *Escherichia coli* XLI-blue competent cells resulted in colonies containing recombinant plasmids with inserts which were analyzed by conventional techniques [14]. Nine clones were picked at random and sequenced by the dideoxy chain-termination method.

Bioinformatics resources

DNA sequences of randomly picked clones were compared with known sequences from GENBANK by the BLAST program (http:// www.ncbi.nlm.nih.gov/BLAST/), IDENTIFY Database (http://motif.stanford.edu/identify), PROSITE Database (http:// www.expasy.ch/spot/prosite/html), The Bio-Web Database (http://molbiol-tools.ca/), S/MARt Database (http: // transfac.gbf.de/SMARtDB/), program OMIGA. The sequence-depending DNA curvature was calculated using the http://lfd.uiuc.edu/staff/gohlke/curve/ program.

RESULTS AND DISCUSSION

This study aims to further characterize tightly DNA-bound complexes from eukaryotic cells. DNA–polypeptide complexes are not completely dissociated during nuclease digestion. After a prolonged digestion of MEL, DNA with DNase I, 0.05–0.1% of undigested DNA can be recovered, implying protection of DNA by the associated proteins. The fraction of protected DNA comprises DNA fragments of different lengths. The main fraction of DNA fragments has an approximate length of 200–300 bp (not shown). Following deproteinization of the nuclease-released DNA–protein complexes, residual blunt-ended DNA strands were cloned into SmaI site of pUC19 vector (see Materials and Methods). Twenty of 40 randomly selected plasmids showed inserts longer than 100 bp. Nine clones of the transformants were picked at random and sequenced. Figure 1. represents the sequences of these DNA inserts. Some properties of sequenced DNA are summarized in Table 1.

The southern blot hybridization of total murine DNA with sequences 1–9 showed the presence of these sequences in middle-copy numbers (results not shown). No significant similarities were found among the different DNA clones. Thus, the nucleotide sequence itself seems to provide only limited information on the nature of DNA from TBD. Recently a degenerate bipartite sequence MRS (MRS : AATAAYAA and AWWRTAANNWWGNNNC, where Y = C or T; W=A or T; R = G or A; N = G, A, T or C) has been identified, which by 80% predicts S/MARs [15]. The bipartite consensus sequences characteristic of S/MARs were absent in all clones studied. Only a small sequence element

Table. **Properties of cloned DNA**

Sequence	Length (bp)	TA + CA + TG content (%)	Alu rep consensus sequences	Topoisomerase II recognition sites
Seq-1	333	18.9	33	6
Seq-2	229	20.6	24	5
Seq-3	161	18.1	5	2
Seq-4	253	32.9	41	12
Seq-5	203	20.7	16	4
Seq-6	385	23.0	29	10
Seq-7	147	27.4	26	7
Seq-8	373	25.4	46	8
Seq-9	425	29.0	29	17

1-seq

ATTAAAATATTGAATGAAGCTGCACAGAAGCAGATCCATGCCCGGGAGAGAGAATAGAGTAAA
GATACTTACTGTAACTGTGCCAGGGGAGATGGGCCAGAAGACAAAAAATGGCTGGCGGTTTTGTGTGGCACC
CCCCCCCCCCAACCCCCAGGAAAAATGCCTTTAAGCCAAGCTTGGTTTATTTCTTTAACAGTAAT
TCATTTCCATTCTCTACAAAGTTGTTAGCTCTTGGGTAAATACCATGTCTTTTTGTATCCTTTTTAAATGTGG
ATTTCGAGCCTGTTAACCCATTTACCCTTGGGCTCTTCCTCAAGATAGATAATCTTGGAG

2-seq

CTCGGTACCCATTAAAATATTGAATGAAGCTGCACAGAAGCAAGTCCATGCCCGG
GTAGAGAGAACAGAGTAAAGATACTTACTGTAACTGTGCCAGGT
GTGAGATGGGCCAG**AATAACAA**AAAATGGCTGGCTTGGTTTTGTGTGGCACTCTCCCC
CCCCCCCAACCCCCAGGAAAAATGCTTTTAAGCAAGTTGTTTATTTCTTTAACAGTA
ATTCATTTCATTCTC

3-seq

ATTAAAGCATATTGACTGCACAGAAGCAGATCCATGGCCCCGAGACGACGAATAGAGTAAAGATA
CTTACTGTAACTGTGCCAGGGGAGATGGGCCAGAAAAGACAAAAAACTGGCTGGCGGTTTT
GTGTGGCACCCCCCCCCCCTTCCCCCAGGAAAAAT

4-seq

CATTTTTGAGATATTAAAAGGTGATATCCAAAGGTTGATTTTACACACACACACACACACACA
CACACACACACCCACACACACACACACATATATGCATACACGTTGCAGGC
ATATGTGTAGAATATATGTATGCATGAACACTTGTGGAGTTGTTGAGCAG
GTGTTAGGCCACAAGTGTGTGTCTGTATGTGTCTATGTGTATGCTTGATTCTCATTTATTAC
TCTTCTTTCATTCTGCTACAAAATTTATGCC

5-seq

CGGCCAGTGAATTGCTCGGTACCGTAGTAACACAGGTCACTGTAAGAACATGACTATTACGTT
TATCTTGAATGGCTTCTGTTAGAAGTCTGGTACCCAGCTTGAAACCACGAGGCAGTGATATCTGAAT
CGGGGAGGGATTTTCTGGAATCTATGAGTCATTTAGGGTCTATATAGCTCCACTATTTCATTTTCACTTCTGA

6-seq

TAAGAAATGAGCAGCGAATCCAGTCAAGAACCCGAGATGGTAAGGAAT
TATATATTGTGTTTTCCTTGCAATCTGTTTTTCCAG**AATAATAA**GAATTGCTATAGTTTGGAT
ATGACTTAAGTGTGTCCCTAGGATGAGATAGGGAGGAAGGATGCTCTCTTTAAAGGTAGGGCC
TTTGGGACATGGCTGGACTAGATAAGGTCATCAGGGCGGACGTCCCTGTGGAT
GAGTAGTGATGGCATTGTAACAAGAGGGGGAGATGGGCACACATGACAT
ACACATATACATATAAATACACCCATACACATGCATGCAGGCATGTGCGTCCATACCCCTTGTCT
CCTTGTCCTGGACTGACCCAGGACTTCCAGTGAAAGACCACCAG

7-seq

TTAGTAAGTTGATGCTGATGGTTGACTCTAGA
GGATTCCTTGGTTGGTTTTTAATTTGGAAGCAAGGAGACGGGGTATGGACG
GCGTGCCTGCATGCATGTGCTATGGGGTGTATTTTATATGTATATGTGTATGTC
ATGTGTGCCA

8-seq

CAGCAACCACATGGTGGGTCACAACCACCCATAATGAGATCTGACTCCCTCTTCTGGAG
TGTCTGAAGTCAGCTCTGGTGTACTTACATATAATCAATAAATAAATCTTAAAAAAAAAAAGTCAGGTAA
GTAGTAGCTCTTTGCTTACTGAGAACAATACCATGATTGCAAGATATTAC
TTCGAATAGTTTTACATATTTTTTAAAATTTCACTAGATAATTAAGACT
TTGAAACATGTATATATTCGTACATATATATATC
GTGTGTGTAATTTTGTCACCTATGACTTAGCTTATAAAAACATAGCCTTTTCAGAG
TAGGTTGATACGATCTGAATACCAAAATGATACTGGTGTTTCAC
ATGTGGTCACAG

9-seq

ATGGAATCGTCGACATCGACGTGACATACTGTATGCATGACAGTCTACGCAACTATA
GAACATACAGAAGTCATCACATATTCTCTCTTGATTCGATTCATATATCATCAGATCAATCATCCG
ACTCACATTACACATAAGCAATCTCAATCTCTCTCATCATTTTCATTGATCGTTATATTCATGATATCG
AATTCATTCTGTCTGTTACTCCATCAAGTAATGATCGACATTCATCATGAATATATATA
AACATACACATATGACATGACACAGTATATATAGTGTCATGGTTATCTATCGATCG
TAGCTATACTGCTACTAGATCGCATAACGTCACGCCATGGACAAATGGCTATATCAGATC
ATATAAGATCGCGATGCATAGATCGATCAATCGATCGATCGACCGAGAGGGCATGCAT

Fig. 1. Nucleotide composition of cloned MEL sequences tightly associated with proteins. AATAAYAA (Y = C or T) sequence elements proposed for S/MAR concensus are bolded

was found in sequences 2 and 6. These elements are bolded in Fig.1. Therefore it can be concluded that DNA present in TBDs is not typical S/MARs. However, due to limited information (only nine sequenced clones) this conclusion is not strict.

However, some other S/MAR properties are shared in these sequences. DNA replication is known to be associated with nuclear matrix attachment sites [2, 3, 12]. The origins of replication share the ATTA, ATTTA and ATTTTA motifs. These motifs are abundant in S/MARs [2] and are presented in most of sequenced DNA clones (Fig. 1). This implies that TBD complexes (as well as the nuclear matrix) can be involved in some development processes.

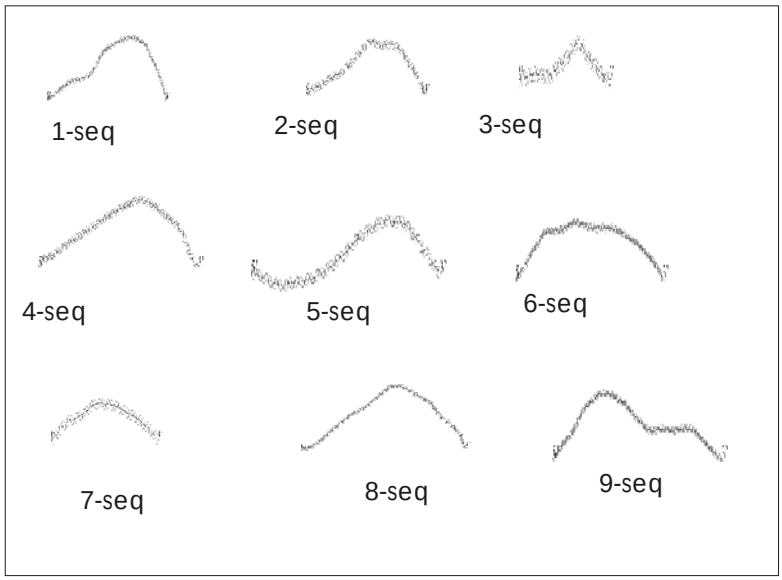

Fig. 2. Conformation of the DNA sequences presented in MEL TBDs. A two-dimensional display of 3-D coordinates

Alu repeats are the largest class among the short interspersed repeated DNA elements (SINEs) found in genomes of higher eukaryotes [16]. They are about 300 bp long and comprise 3–6% of the haploid genome. Alu repeats are mobile regulators of transcription, which are able to switch on and off certain genes [16]. Transcripts from Alu repeats form stable Alu-RNP particles. Their function remains unknown. It has been suggested that they take part in the regulation of protein translation [17]. As seen in Table 1, all randomly picked clones contain Alu repeats. These data are consistent with data of other authors who assume that most of DNA fragments tightly bound to proteins belong to the Alu family of sequences [10, 18].

All S/MARs are to have numerous topoisomerase II cleavage and binding sites [2, 12]. This enzyme can lead to relaxation of both positive and negative supercoils. These observations suggested the possibility that the nuclear matrix might be involved in quite diverse processes, and all of them contribute to changes of the topology of chromatin domains. As seen in Table 1, all cloned sequences have numerous copies of the sequence elemens which are shown is consensus binding sites for topoisomerase II [19]. Therefore the reversible formation of complexes between specific DNA sequences and proteins in the nucleus might provide quite a general means of integrating TBD structure and function [10, 11]. This idea supports the observation that TBD DNA fragments contain multiple copies of sequences related to the topo II concensus. S/MARs are thought to flank both sides of genes, so TBD complexes may define the ends of topologically isolated matrix-bound domains in the chromosome [7, 8].

It has been shown that CA, TA and TG dinucleotides are overrepresented in DNA sequences which are nuclear protein recognition sites [20]. Some (not all) S/MARs display an unusual richeness in TA, TG and CA [2]. As seen in Table 1, most of DNA sequences from TBD complexes contain more of CA, TA and TG as compared to randomly distributed sequences (in randomly distributed sequences CA, TA and TG content is 18.75%). Computer analysis has also revealed that cloned DNA sequences contain multiple concensus motifs, which are recognized by known transcription factors (results not shown). Therefore DNA of TBD complexes can provide a means of bringing together the components of the transcription system. It is of interest to note that computer search revealed three of the DNA sequences studied to contain Z-DNA stretches (seq-4, -6 and -9). Sequence-4 contains 17 Z-DNA stretches. Left-handed Z-DNA segments can serve as regulatory elements during DNA replication, reparation, etc. Therefore TBD complexes may be involved in the regulation of gene expression.

Curved DNA is thought to play an important role in nuclear processes that involve the interaction between DNA and proteins. Curved DNA has been identified at or near the sites where S/MARs are attached to the nuclear matrix [21]. Recently, on the basis of the DNA unwinding potential of DNA sequences, there have been developed computer programs that calculate the DNA unwinding potential of a given sequence [22]. Applying these programs to cloned DNA sequences we have found revealed that these sequences contain the regions of a strong bending potential. Results are presented in Fig. 2. It can be considered that all DNA studied consist of strongly curved regions. These results coincide with the properties of S/MARs [21]. These data suggest also that DNAs presented in tightly DNA-bound complexes have a characteristic conformation instead of consencus sequences.

The structure and function of TBD complexes are far from being clear. The earlier reported results differ and sometimes contradict each other. In this study, we try to compare the properties of sequence-based DNA from TBD with known S/MARs properties. Here we demonstrate a correlation between some properties of DNA sequences involved in TBD complexes and sequences identified as S/MARs. The identification of proteins that bind to DNA in TBD complexes will be the next step towards the elucidation of the possible role of TBDs.

This work was partly supported by the European Commission, Program of Centres of Excellence, project CEBIOLA, ICA1-CT-2000-70027.

Received 11 February 2004
Accepted 29 November 2004

References

1. Lamond AI, Earnshaw WC. Science 1998; 280: 547–53.
2. Boulikas T. J Cell Biochem 1993; 52: 14–22.
3. Paulson JR, Laemmli UK. Cell 1977; 12: 817–28.
4. Fernandez LA, Winkler M, Grosschedi R. Mol Cell Biol 2001; 21: 196–208.
5. Shera KA, Shera CA, McDougall JK. J Virol 2001; 75: 12339–46.
6. Will K, Warnecke G, Albrechtsen N et al. J Cell Biochem 1998; 69: 260–70.
7. Tsanev R, Avramova Z. Chromosoma 1994; 103: 293–301.
8. Razin SV, Hancock R, Jarovaia O et al. Cold Spring Harbor Symp on Quant Biol 1993; LVIII: 25–35.
9. Juodka B, Pfutz M, Werner D. Nucl Acids Res 1991; 19: 6391–8.
10. Neuer-Nitsche B, Lu X, Werner D. Nucl Acids Res 1988; 16: 8351–60.
11. Avramova Z, Georgiev O, Tsanev R. DNA Cell Biol 1994; 13: 539–48.
12. Jack RS, Eggert H. Eur J Biochem 1992; 209: 503–9.
13. Labeikyte D, Bagdoniene L, Juodka B. Biologija 2000; N2 (priedas): 88–91.
14. Current Protocols in Molecular Biology, John Wiley & Sons, Inc. 1994.
15. van Drunen CM, Sewalt RGAB, Oosterling RW et al. Nucl Acids Res 1999; 27: 2924–30.
16. Calos MP, Miller GN. Cell 1980; 20: 579–85.
17. Bovia F, Strub KJ. Cell Sci 1996; 109: 2601–08.
18. Werner D, Neuer-Nitsche B. Nucl Acids Res 1989; 17: 6005–15.
19. Sander M, Hsieh T-S. Nucl Acids Res 1985; 13: 1057–71.
20. Trifonov EN . Trends Biochem Sci 1991; 16: 467–70.
21. Yamamura J, Nomura K. FEBS Letters 2001; 489: 166–70.
22. Shpigelman ES, Trifonov EN, Bolshoy A. Comp Apppl Biosci 1993; 9: 435–40.

D. Labeikytė, L. Bagdonienė, V. Jonuðienė, S. Sasnauskienė, B. Juodka

PELIØ DNR BALTYMINIØ KOMPLEKSØ DNR SEKØ SAVYBIØ PALYGINIMAS SU S/MAR

S a n t r a u k a

Atlikta DNR fragmentø, iðskirtø ið stipriai su DNR susijusiø DNR baltyminiø kompleksø sekø, palyginamoji analizë. Devyniø nepriklausomai iðskirtø ir sekvenuotø klonø kompiuterinë analizë parodë, kad ðios sekos pasiþymi kai kuriomis (ori signalai, TA, CA, TG dinukleotidø gausa, topoizomerazës II atpaþinimo vietos, DNR linkio vietos) S/MAR sekø savybëmis. Tai rodo, kad DNR fragmentai, iðskirti ið stipriai su DNR susijusiø DNR baltyminiø kompleksø, pasiþymi specifinëmis konformacinëmis ir sekos S/MARs savybëmis.